Research on personalized recommendation system to improve overall diversity

Liyan Liu[1]

[1]Shandong University of Political Science and Law, Ji Nan, Shandong Province, China

Abstract. The recent research on the impact of information technology on consumers' purchasing behavior shows that the purchasing power of consumers who buy more diversified products is stronger than that of consumers who buy single types of goods. Improving the overall diversity of recommendation results can make more products appear in the user's recommendation list.

1 Introduction

For each user, it is very difficult but very important to identify the useful information from the massive information. Personalized recommendation system is an important method to solve this problem. It helps users select the most appropriate information from the huge data set. By determining the user's preferences[1], the recommendation system recommends the most suitable or interesting items to a specific user. Recommendation algorithms mainly include content-based recommendation, collaborative filtering recommendation and hybrid recommendation. Recommendation system is widely used in movies, music, books, tourism, e-commerce, social networking and network search. Accuracy is an important indicator to evaluate the recommendation system. It is to evaluate whether the items recommended to users are the most suitable. However, the information users recommended in this way may have been obtained from other channels, so it is unnecessary in many cases. The diversity of another important index to evaluate the advantages and disadvantages of the recommendation system has attracted more and more attention of researchers and users. The diversity reflects the difference of recommended items. Some scholars even claim that diversity in some cases exceeds precision performance and brings better satisfaction to users.

2 The present situation, level and development trend of technology at home and abroad

The accuracy and diversity of recommendation system are two completely different aspects. A good recommendation system should take into account these two recommendation standards at the same time. However, these two standards are mutually restricted. The increase of accuracy will inevitably lead to the decrease of diversity, and the increase of diversity will also lead to the decline of accuracy. How

to find a balance between the two is an urgent problem The problem. Collaborative filtering recommendation system makes recommendations based on the user's rating data. However, there are often problems such as uneven number and sparsity in the rating data set, which results in errors in the recommendation results. This kind of error has a great impact on the accuracy and diversity of recommendation. At present, a lot of research is focused on the optimization accuracy, rarely involving diversity, especially the overall diversity Sex. How to improve the overall diversity of recommendation results is another important problem to be solved. Effectively improving the overall diversity of recommendation results can not only meet the personalized experience requirements of users, but also improve the effect of long tail marketing, so as to help enterprises maximize profits.

2.1 Research status of recommendation algorithm

Common recommendation algorithms include content-based recommendation algorithm, collaborative filtering recommendation algorithm, hybrid recommendation algorithm, and other recommendation algorithms, such as: utility based recommendation, knowledge-based recommendation, association rule-based recommendation. No matter the traditional algorithm or the improved algorithm, the ultimate goal is to recommend the most suitable products to customers. Most of the recommendation systems use the accuracy to evaluate the recommendation algorithm. But, purchasing behavior shows that the purchasing power of consumers who buy more diversified products is stronger than that of consumers who buy single types of goods. Therefore, in recent years, researchers of recommendation system have gradually shifted more attention from the traditional recommendation strategy based on accuracy to the diversity of products. The products recommended to users are not only limited to the best-selling products

Liuliyansd@163.com

recommended by the accurate recommendation, but also often contain some new and strange niche products.

2.2 Research status of recommended diversity

At present, the definition of diversity is divided into two categories: individual diversity and overall diversity. Individual diversity is a measurement standard from the perspective of a single user. The goal is to recommend some products with low similarity to each other but in line with the user's interests for a specific user. The overall diversity can be understood as the number of different items recommended to different users. Overall diversity is not directly related to individual diversity, it reflects the ability of recommender system to recommend different items to different users. However, in many cases, although the traditional recommendation method can get high accuracy recommendation results, but the overall diversity will be affected. Adomavicius et al. Proposed an improved project ranking technology, which can be combined with different rating prediction algorithms to obtain better overall diversity. D.Fleder et al. Studied the impact of recommendation systems on sales diversity. The results show that even some well-known recommendation systems may lead to the reduction of sales diversity, because these systems recommend products on the basis of sales and ratings. T. Bobadilla et al. Proposed several optimization based methods to improve the overall diversity, including greedy algorithm, maximum flow based method and integer programming method. However, the balance between accuracy and diversity is not ideal.

2.3 Research status of similarity computing

Computing the similarity between users is an important part of collaborative filtering algorithm. There are many methods to calculate the similarity, including person correlation coefficient similarity, Jaccard similarity, cosine similarity and modified cosine similarity. Wang Cheng et al. Proposed a modified Jaccard similarity coefficient to calculate the similarity between users, so as to obtain a more accurate user similarity matrix[2]. Wang Fuqiang et al. Proposed an improved cosine similarity calculation method[3], which established a project user backlookup table, only calculated the similarity between users with the same rating items, which avoided the huge workload of calculating the similarity between two users in the traditional method. Zhu Jialong and others proposed a new asymmetric person similarity algorithm to improve the accuracy of recommendation[4]. All the methods mentioned above are based on the premise of improving the accuracy of recommendation, but have no obvious effect on the recommended diversity, especially the overall diversity.

3 Main problems

The current recommendation system has achieved good results in customer satisfaction and helping enterprises to make profits. However, there are still some problems in the following aspects. ① Most of the existing recommendation technologies emphasize on improving the

accuracy of the recommendation content, while ignoring the diversity of the recommendation results, which leads to the homogenization of the recommendation results, reduces the customer's experience satisfaction, and is not conducive to the sales of long tail products. ② Recommendation accuracy and diversity are two mutually restricted indicators of evaluation and recommendation system. The improvement of diversity is usually based on the loss of recommendation accuracy. How to improve the overall diversity of recommendation results while ensuring the accuracy is another important issue.③ User similarity calculation ignores the difference in the number of user ratings, and does not consider the impact of the number of common scoring items on user similarity calculation. Moreover, the existing similarity calculation methods are aimed at improving the accuracy of recommendation, ignoring the recommendation diversity, especially the overall diversity. ④ In the traditional collaborative filtering algorithm, the influence of the number of items jointly scored by users on the similarity calculation is not considered in the process of item prediction and scoring, but all the scored items are given the same weight, resulting in errors in the recommendation results.

4 Research content

4.1 The relative similarity index algorithm is studied

In many collaborative filtering recommendation systems, similarity calculation methods are often used, including cosine similarity, person correlation coefficient, Jaccard similarity, etcetera. It has been proved that these methods are very successful in many researches. However, it is found that in many cases, when the above similarity calculation methods are used to improve the accuracy, the overall diversity of the recommendation results is negligibleIt doesn't help.The research content of the project is to refer to the number of common scoring items among users, and the impact of users' score difference on user similarity in similarity calculation, and propose and implement an optimized similarity calculation method relative similarity index algorithm, which provides accurate guarantee for obtaining more accurate prediction score[5].

4.2 This paper studies the prediction scoring method based on relative similarity index

In the collaborative filtering recommendation system based on users, the purpose of similarity calculation is to calculate the user's nearest neighbor, and then use the nearest neighbor to predict and score items. The result of prediction score is directly related to the content of the recommendation list. The project plans to refer to the relative similarity index to revise the prediction scoring method. By adjusting the weighted common scoring items, the calculation weight of more common scoring items will be increased, while the calculation weight of less common

scoring items will be reduced, so as to improve the accuracy of prediction scoring.

4.3 This paper studies the collaborative filtering recommendation model of overall diversity optimization s

The research of individual diversity in personalized recommendation has made great progress compared with previous years, but the research on overall diversity is relatively less. This project intends to study the prediction scoring algorithm based on relative similarity calculation and correction, and the collaborative filtering recommendation model with overall diversity optimization based on the reference score threshold, which can ensure the accuracy of recommendation and improve the overall number of recommendation resultsIt's not.

4.4 This paper studies the application of recommendation model in enterprises recommendation model of overall diversity optimizations

The project intends to deeply understand the relationship between long tail marketing and recommendation diversity through literature research, and verify the impact of overall diversity on the sales of long tail products through open data set test, optimize the recommendation model, cooperate with e-commerce enterprises, test the impact of recommendation system on long tail marketing and long-term profit, formulate reasonable strategies and improve the application effect.

5 Main technologies

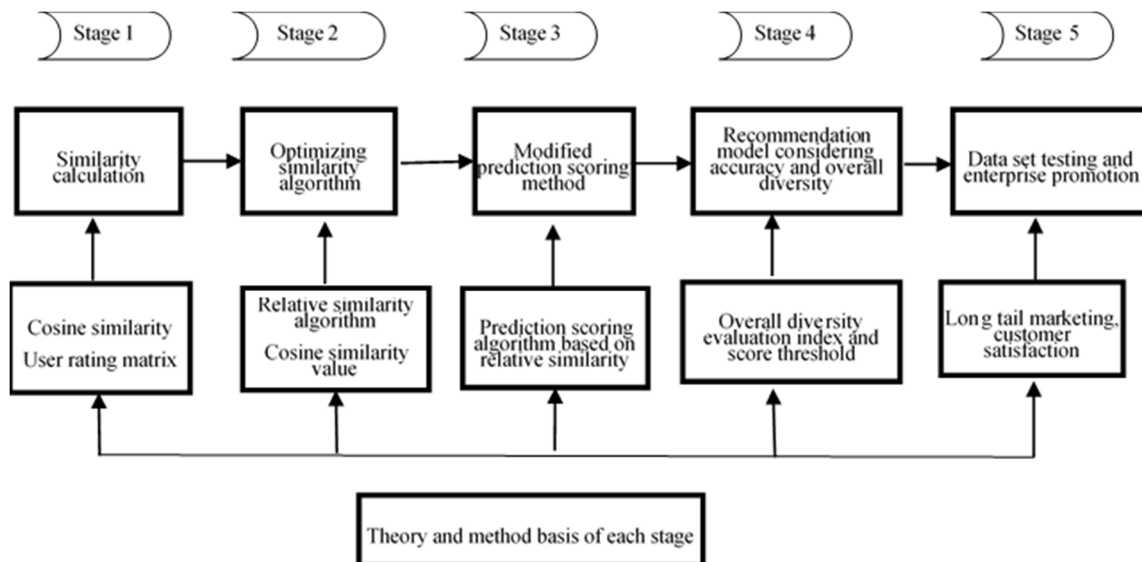The research process and technology are shown in Figure 1

Fig1. Technical route and research steps

6 Expected problems and Solutions

6.1 Similarity calculation

For example, if there are four users and related items, Ln (u1) = {i1, i3, i5} , Ln (u2) = {i3, i5, i7} , Ln(u3)={i3} and Ln(u4)={i2, i3, i7}, in this case, it is generally considered that the cosine similarity of two individuals is 1 no matter how different their scores are, and u3 is regarded as the nearest neighbor of the other three users. Such similarity calculation will make the prediction score appear larger error, which will cause great misleading to users.

The reason for the above problem is that the number of common ratings of two users is not considered. In reality, the more items two users jointly score, the higher the similarity is. Therefore, the number of items shared by users should be considered as an important factor in similarity calculation. The method to solve this problem is to add a weight to deal with it.

6.2 Evaluation index of overall diversity

Recommended diversity includes individual diversity and overall diversity, and their evaluation indexes are also different. Usually, the method to evaluate individual diversity is to calculate the similarity among the items in the recommended list. In this project, the overall diversity optimization is to be realized. How to evaluate the optimization performance of the recommendation model is an important problem faced by the project. The evaluation index to be used in the project is Hamming distance. Hamming distance is used to measure the difference between the two user recommendation lists, such as

formula (1) $\quad H_{ut}(L) = 1 - \dfrac{Q_{ut}(L)}{L}$ .Where Q_{UT} (L)

represents the number of the same products in the recommended list of users u and t.If the two recommended lists are identical, then H_{ut} (L) = 0; otherwise, if the two recommended lists have no overlapping products, then H_{ut} (L) = 1.The result of the above formula is the

recommended dissimilarity between any two users, and then the average Hamming distance between all users is the overall diversity evaluation index of the whole system. The higher the value, the higher the diversity of recommendations.t

7 conclusion

The traditional recommendation algorithm often recommends some popular products, which generally have a high prediction score value. However, some unpopular products in the long tail part are difficult to appear in the recommendation list because most users do not know in advance. In fact, these goods are likely to meet the interests of users. Some studies use the long tail effect to explore the impact of the overall recommendation diversity on the sales of goods. It is found that the recommendation system with higher diversity is conducive to promoting the sales of less popular products. Other studies have shown that improving the overall diversity of the recommendation system also has many benefits for users, so that the system re recommends products that can provide users with a wider range of choices, rather than the best-selling products that users can find themselves. This paper proposes and designs a relative similarity index algorithm to improve the prediction scoring method, designs and implements a collaborative filtering recommendation model to optimize the overall diversity of recommendation results under the premise of limited accuracy, so as to improve the user's experience satisfaction, improve the long tail marketing effect of commodities, and help enterprises maximize profits.

Acknowledgment

References

1. Adomavicius G. and K won Y, A Treatise on Electricity and Magnetism, Optimization-based approaches for maximizing aggregate recommendation diversity[J], INFORMS Journal on Computing.2019, (24) :896-911. [2] Nunez-Valdez E. R., Cueva-Lovelle .T.M. , Sanjuan-Martt'nez 0., Garct ' a-Dt'az V., Ordonez P., Montenegro-Man' n C.E., (2012)Implicit feedback techniques on recommender systems applied toelectronic books, Computers in Human Behavior.2018, 28 (4):1186-1193.

2. Wang Cheng, Zhu Zhigang, Zhang Yuxia and Su Fangfang, "Improvement of recommendation efficiency and personalization of user based collaborative filtering algorithm," Minicomputer system, 2016(03):428-432

3. Wang Fuqiang, Peng Furong, Ding Xiaohuan and Lu Jianfeng, "Collaborative filtering recommendation algorithm based on location asymmetric similarity measure," J Computer application., 2016, 36(01):171-174.

4. Zhu Jialong, LV Yuqin and Li Weihai, "A new collaborative filtering algorithm based on asymmetric Pearson similarity," [EB/OL] Beijing: China Scientific Papers Online http://www.paper.edu.cn/releasepaper/content/,2014 01-394.

5. Jiang Shuhao, Zhang Liyi and Zhou Na, "Collaborative filtering recommendation based on dynamic change of user preferences" [J] , Computer and modernization, 2020(1).